

## МОДЕЛИ В ЭВОЛЮЦИОННОЙ БИОЛОГИИ

Н.Н. Юрченко, И.В. Дейнеко, И.К. Захаров

Учреждение Российской академии наук Институт цитологии и генетики Сибирского отделения РАН, Новосибирск, Россия, e-mail: yurchen@bionet.nsc.ru

Дается описание математических подходов к эволюционной биологии. Рассматриваются общие принципы, цели и стратегия математического моделирования в биологии от работ биометриков, современников Ч. Дарвина, математической формализации микроэволюционных процессов – влияния мутирования, миграции, изоляции, дрейфа генов на генофонд популяций до современного этапа филогенетических реконструкций на основании секвенированных последовательностей ДНК.

**Ключевые слова:** эволюция, популяции, генетическая изменчивость, адаптация, филогенетическая реконструкция, математические модели.

Математические модели в биологии, и в частности в эволюционной биологии, являются неотъемлемой частью современных научных исследований, и чем глубже становится понимание некоторого явления, тем детальнее и сложнее становится его математическая модель (Фишер, 1958; Smith, 1968; Ратнер, 1977; Математические модели ..., 1980; Свирежев, Пасеков, 1982; Животовский, 1984; Кель и др., 1989; Колчанов, Шиндялов, 1991; Seger, Adler, 2002). И хотя в фундаментальной работе Ч. Дарвина «Происхождение видов ...» (On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life, 1859) (Дарвин, 1939) нет формул, матриц или графов филогенетических связей, при рассуждениях о динамике идеализированных биологических ситуаций, тем не менее, автор смело оперирует количественными понятиями, которые затем были формализованы его последователями в математические формулы и модели.

К настоящему времени в дополнение к эволюционному вербальному мышлению создан широкий набор математических моделей, использующих изощренные математические методы, численные результаты которых получают на суперкомпьютерах, однако цели и общая стратегия моделирования остаются прежними: строгое постулирование набора допущений и логических правил и вывод на их основе

правдоподобных заключений. Математическое моделирование эволюционного процесса в идеале должно дать ответ на основной вопрос эволюции: могут ли случайные, или, образно выражаясь, слепые процессы изменчивости, наследования и естественного отбора служить исчерпывающим основанием для объяснения прогрессирующего в историческом развитии адаптогенеза и наблюдаемого в природе огромного биоразнообразия?

Ч. Дарвин и его современники понимали, что ответ на этот вопрос кроется в числах, вероятностях и изменчивости частот признаков во времени и в пространстве, но не могли ответить на него, в частности потому, что ничего не знали о природе генов как об основе эволюционного процесса и материальных носителях наследственной информации. И только зарождение генетики в самом начале XX в. с переоткрытием законов Г. Менделя подвело естественнонаучный базис под объяснение сущности наследования и изменчивости биологических признаков. Это открыло дорогу для математического моделирования генетических явлений и теоретического конструирования эволюционного процесса.

Отметим главные начальные вехи этого пути:

1) в 1908 г. Г. Харди и В. Вайнбергом было независимо показано, что в диплоидных популяциях, для которых постулируется случайное

скрещивание, различные генотипы по данному гену будут появляться снова и снова, из поколения в поколение в прежней пропорции (Provine, 1971; Кайданов, 1996; Gillespie, 1998; Алтухов, 2003);

2) через десять лет были заложены основы теоретической популяционной генетики Р. Фишером, Дж. Холдейном и С. Райтом, их математические построения показали, что синтез теории естественного отбора с менделизмом подтверждает исходные теоретические дарвиновские построения. При этом их модели создали прочный генетический базис для оценки динамики количественных признаков в популяциях (Fisher, 1930; Haldane, 1932);

3) модель структуры ДНК, предложенная Дж. Уотсоном и Ф. Криком в 1953 г., и последующее развитие методов расшифровки первичных последовательностей ДНК дали толчок методам кладистики, основанным на теории графов. Молекулярные генетики открыли богатейшие источники эволюционной информации в последовательностях ДНК, РНК и белков, а систематики научились использовать эту информацию для установления филогенетического родства между популяциями, видами и высшими таксонами (Sokal, Sneath, 1963; Cavalli-Sforza, Edwards, 1964, 1967; Hinnig, 1966; Felsenstein, 1983; Ridley, 1986; Hull, 1988; Harvey, Pagel, 1991; Pagel, 1994; Swofford *et al.*, 1996; Kitching *et al.*, 1998; Larget, Simon, 1999; Lewis, 2001; Lutzoni *et al.*, 2001; Погозин и др., 2005; Шедько и др., 2007);

4) со второй половины XX в. эволюционное моделирование становится более разнообразным и многоплановым: экологи и этологи нащупывают подходы к формулированию общих принципов управления эволюционно-историческим развитием экосистем и биоценозов, к объяснению поведения индивидуумов в популяциях и формированию социально-иерархических структур с использованием математической теории игр;

5) широкое внедрение микрочипов и технологий для анализа продуктов транскрипции в клетках привело в последнее десятилетие к развитию новых статистических методов (Ефимов, Катохин, 2009), а также теории построения генетических сетей (Лихошвай и др., 2003).

При этом надо подчеркнуть, что хотя математическое моделирование играет ключевую роль

во всех направлениях развития современного биологического познания, только сравнительно небольшая часть биологов вносит вклад в развитие собственно математических моделей и методов моделирования или хотя бы имеет представление о масштабах современных биоматематических исследований.

Далее мы обсудим математические модели и их вклад в развитие трех важнейших областей эволюционной теории, соответствующих основным элементам дарвиновской аргументации: 1) модели, характеризующие генетическую изменчивость внутри популяций; 2) модели, объясняющие изменение популяции во времени и пространстве и образование новых видов; 3) модели, анализирующие адаптивную эволюцию фенотипов конкретного вида.

### Как работают эволюционные модели?

Кроме математических моделей есть и простые, условно назовем их «рукотворными». Например, модель двойной спирали молекулы ДНК Дж. Уотсона и Ф. Крика, которая включает в себе конкретную гипотезу о химической структуре ДНК.

Моделью можно назвать любой формальный метод, который начинается с рассуждения «что если» и заканчивается выводами «тогда». Достаточно наглядной является широко известная модель взаимодействия двух популяций – хищника и жертвы, описываемая уравнением Лотка–Вольтерра (Lotka, 1925; Вольтерра, 1976, Родин, 1991). Численность популяции жертвы  $N_1$  будет изменяться во времени (зависит также от численности популяции хищника  $N_2$ ) согласно уравнению:

$$\frac{dN_1}{dt} = r_1 N_1 - p_1 N_1 N_2,$$

где  $r_1$  – скорость размножения популяции жертвы,  $p_1$  – коэффициент хищничества для жертвы (вероятность того, что при встрече с хищником жертва будет съедена). Прирост популяции хищника описывается уравнением:

$$\frac{dN_2}{dt} = p_2 N_1 N_2 - d_2 N_2,$$

где  $d_2$  – смертность хищника,  $p_2$  – коэффициент хищничества (некая величина, указывающая на выгоду, полученную хищником при поедании жертвы).

Решением данной системы являются гармонические колебания, в частных случаях – с затухающей или увеличивающейся амплитудой. Динамика численности хищника и жертвы, описываемая данной моделью, не является структурно устойчивой, т. е. малое изменение параметров модели (например учитывающее ограниченность ресурсов) может привести к качественному изменению поведения и даже вырождению (Базыкин и др., 1980). Однако простота и наглядность положенных в основу этой модели предположений и правдоподобность результатов сделали ее столь популярной.

Такие модели позволяют проникнуть в суть высказываемой гипотезы путем визуализации ее основного смысла. Однако часто бывает так, что если гипотеза подтверждается эмпирическими данными, то она попадает в разряд занятых научно-исторических артефактов, хотя базисные предположения могут быть и неверны. Так, согласно законам математической логики (Клини, 1973), из двух ложных предположений можно вывести истину, например: камень съедобен; камень есть хлеб; следовательно, хлеб съедобен (Ортега-и-Гасет, 1991). Так, в вышеприведенной модели «хищник–жертва» циклические колебания численностей некоторых популяций могут быть предопределены, например, сменой времен года или распространением пандемий и не иметь ничего общего с математически описанным выше взаимоотношением хищника с жертвой. Кроме того, отметим, что при моделировании некоторого феномена правильно было бы говорить об описании моделью его поведения, а не о вскрытии сущности *per se* исследуемого явления. Так, открытие законов всемирного тяготения И. Ньютоном ровным счетом ничего не дает для объяснения феномена гравитации: «Причину же этих свойств силы тяжести я до сих пор не мог вывести из явлений, гипотез же я не измышляю... Довольно того, что тяготение на самом деле существует и действует согласно изложенным нами законам и вполне достаточно для объяснения всех движений небесных тел и моря» (цит. по: Клайн, 1988).

Скептики могут сказать, что модели в эволюционной биологии часто являются гипотезами, которые не могут быть непосредственно проверены экспериментально, во-первых, из-за временных ограничений, накладываемых на них,

во-вторых, из-за недостаточного соответствия теоретических моделей реальности, в-третьих, многие эволюционные модели являются «ложными» в важных отношениях, прежде всего потому, что типичная модель намеренно упрощает некоторые сложные аспекты объекта исследования. На самом деле гносеологический вопрос заключается в том, действительно ли мир устроен так, как предполагает модель, и может ли он быть лучше понят при его упрощении и игнорировании большинства параметров определенного процесса за исключением нескольких потенциально существенных? Для установления общих принципов, которые далее могут быть применены к реальности во всей ее многосложности, эволюционные модели обычно имеют дело с небольшим числом переменных (например,  $z$  зависит только от  $x$  и  $y$ ), где  $z$  действительно зависит от сотен переменных. Можно назвать два крайних подхода при моделировании, хотя часто для решения практических задач авторы выбирают некоторую их комбинацию. Первый подход заключается в определении одного–двух важных параметров системы, при этом постулируется, что остальные параметры либо являются константными, либо изменяющимися заведомо медленнее основного параметра – так называемые квазистационарные системы (Ратнер, Шамин, 1980; Дейнеко и др., 2003). При этом модель получается очень наглядной и простой для численного решения и анализа. Очевидно, что данный подход жестко ограничен исходной ситуацией, и при ее изменении модель перестает работать. Противоположная крайность – описание поведения и реального взаимодействия между собой каждого отдельного элемента системы и последующий вывод на этой основе интегральных законов для всей совокупности (Kimura, Ohta, 1974). Отметим, что этот подход, являясь единственным для достаточно широкого круга явлений, чрезвычайно трудоемок для численного решения. Данный подход первоначально был применен для численного моделирования протекания ядерного взрыва в 1960-х гг., до того как были созданы адекватные интегральные математические модели.

Показательным примером зависимости результатов моделирования от исходных предположений, заложенных в основание модели,

может служить задача построения филогенетических деревьев. Актуальность этой задачи подтверждается большим объемом вновь секвенированных ДНК-последовательностей и широким выбором компьютерных программ для построения филогенетических деревьев, их визуализации и анализа (Omland *et al.*, 2008).

Построение филогенетических деревьев начинается с создания матрицы расстояний между изучаемыми объектами, которая отражает количество различий между нуклеотидными или белковыми последовательностями. При этом полученная матрица может удовлетворять или не удовлетворять условию аддитивности, т. е. условию равенства суммы расстояний от А до В и от В к С к расстоянию, вычисленному непосредственно от А до С. В том случае, если матрица расстояний не аддитивна (а таких случаев большинство), построение филогенетического дерева может быть неоднозначным и зависит от конкретного метода построения, что является отдельной областью исследований (Fitch, Margoliash, 1967; Atlas ..., 1972; Fitch, Farris, 1974; Waterman *et al.*, 1977). Однако главной проблемой при построении филогенетических деревьев являются предварительный анализ предполагаемых свойств и характеристик изучаемых последовательностей и априорный выбор соответствующего метода подсчета расстояния. Очевидно, что количество и характер нуклеотидных замен в межгенной части генома будут отличны от замен в кодирующих частях, которые в свою очередь будут отличаться в интронах и экзонах, а последние будут подразделяться на синонимичные и несинонимичные замены, очевидно, с различными частотами.

Таким образом, эволюционные расстояния, вычисленные по различным частям генома, не будут одинаковыми, хотя задача филогении состоит в определении некоторого конкретного момента расхождения видов. Поэтому дерево, построенное «вслепую», в отрыве от физической сущности рассматриваемых данных, можно считать наименее надежным, а обобщать или переносить результаты такого анализа на другие виды было бы и вовсе некорректно. В качестве примера рассмотрим хорошо изученные филогенетические деревья семейств глобинов (Goodman *et al.*, 1971). Скорости фиксации замен в функциональных центрах на порядок ве-

личин ниже, чем во внецентровых участках, при этом большинство составляют замены, не меняющие полярность или размер соответствующей аминокислоты (Ратнер, Кананян, 1975). Следовательно, игнорирование факта почти синонимических замен в функциональных центрах глобинов может привести к искажению филогенетических взаимоотношений.

Большое значение при оценке временных характеристик филогенетических деревьев имеет «эффективный размер» популяции рассматриваемого вида. Эта характеристика отвечает численности скрещивающихся между собой особей, т. е. численности популяций, передающих свою генетическую информацию следующему поколению. В больших популяциях отдельные локусы ДНК могут начать дивергировать задолго до момента нового видообразования, тем самым увеличивая вариацию времени дивергенции, подсчитанной для различных локусов ДНК. Так, например, могут сформироваться два типа аллелей  $A_1$  и  $A_2$  (либо комбинации аллелей  $A_1B_1C_1$  и  $A_2B_2C_2$ ), практически равных между собой по жизнеспособности, но лучших по сравнению с другими аллелями, например с  $A_3$  ( $A_3B_3C_3$ ), следовательно, они станут доминирующими в популяции. При этом фактическое расхождение видов может быть обусловлено совсем иными причинами и произойти намного позже (Takahata *et al.*, 1995).

Большие популяции несут большее генетическое разнообразие, поэтому в них с большей вероятностью могут образоваться новые подвиды. Это ожидание противоречит малому эффективному размеру популяции в момент видообразования. Так, например, время расхождения человека и неандертальца по разным геологическим оценкам составляет от 360 до 600 тыс. лет. Если предположить, что последний общий предок человека и неандертальца жил около 400 тыс. лет назад, то эффективный размер популяции тогда состоял всего из 3 тыс. особей (Green *et al.*, 2006). Верхняя граница в 600 тыс. лет дает еще меньшую оценку – менее 1 тыс. особей. Еще раз отметим, что это число показывает эффективный размер популяции. Реальный же ее размер с учетом возрастного распределения, смертности и т. п. может быть на 1–2 порядка величин выше. При этом общее население Земли, конечно же, не ограничива-

лось 300 тыс. особями: могли быть и другие «кланы» человеко-неандертальцев, которые в ходе эволюции были вытеснены более успешными сородичами.

Вышеприведенные примеры показывают, что имеется большое число факторов, которые могут влиять на вид филогенетических деревьев. Специфические особенности нуклеотидных последовательностей генетического локуса, на основе которого строится дендрограмма, могут кардинальным образом влиять не только на длину ребер дерева, но и на его топологию. Все это приводит к тому, что невозможно указать единственно правильный универсальный рецепт или план действий для всех случаев. Существуют лишь единичные примеры, когда примененный исследователями тот или иной подход можно было экстраполировать на другие задачи. Более того, параметры, которые исследователь выбрал при моделировании как существенные, могут на самом деле не быть таковыми, или же могут иметь место неучтенные, дополнительные эффекты, которые в определенной ситуации становятся определяющими, например моменты глобальных геномных перестроек.

Несмотря на всю сложность задачи моделирования эволюционных процессов, очевидно, что любая, даже очень достоверная модель может быть успешной только в узком смысле или только для частных упрощенных случаев. При этом само моделирование должно начинаться именно с определения ситуации, для исследования которой модель будет предназначена.

Одной из наиболее известных, наглядных и простых моделей эволюционной генетики является модель Харди–Вайнберга, в которой игнорируются все признаки организма, за исключением одного абстрактного, определяемого двумя типами аллелей некоторого гена в диплоидном локусе. Популяция предполагается бесконечно большой, для которой характерен случайный тип скрещиваний и на которую не действует естественный отбор. Один из аллелей (частота встречаемости которого  $p$ ) обозначается как  $A$ , а другой как  $a$  (частота которого  $q = 1 - p$ ). Логика рассуждений такова: если индивиды скрещиваются случайно, независимо от своего генотипа, то разные диплоидные генотипы потомков будут образовываться в том же отношении, что и у родителей: с частотой  $p^2$  будут встречаться

гомозиготы  $AA$ , с частотой  $2pq$  – гетерозиготы  $Aa$ , и  $q^2$  – гомозиготы  $aa$ , что следует из простого математического правила:

$$1 = (p + q)^2 = p^2 + 2pq + q^2,$$

т. е. частоты аллелей будут оставаться неизменными –  $p$  и  $q$  – из поколения в поколение. Таким образом, соотношение распределения разных генотипов будет воспроизводить само себя каждое следующее поколение. Это может казаться очевидным и банальным для многих читателей, но во времена Харди–Вайнберга некоторыми генетиками были опубликованы альтернативные запутанные теории.

Будучи фундаментальным для популяционной генетики, закон Харди–Вайнберга основан на простейших допущениях, описывающих идеальную популяцию. Однако реальные организмы содержат десятки тысяч генетических локусов, а не два; реальные популяции никогда не являются бесконечными по численности; индивиды скрещиваются не случайным образом; сегрегация аллелей во время образования гамет не всегда бывает правильной; нельзя полностью исключить действие отбора (если не на сам локус  $A$ , то на сцепленные с ним локусы) и т. д. Однако постулируемые упрощения фокусируют наше внимание на фундаментальном вопросе; на практике неучитываемые факторы часто оказывают незначительное влияние, и ими можно пренебречь.

Вскрытие генетического полиморфизма в реальных популяциях и изучение микроэволюционного процесса, происходящего в них, показали, что в общем случае не наблюдается точного соответствия характеристик популяций (частоты аллелей, частоты генотипов и т. п.) с законом Харди–Вайнберга. Поэтому были разработаны некоторые подходы к оценке величины такого несоответствия (Никоро и др., 1980). Так, обобщенный закон Харди–Вайнберга дает соотношение аллелей:  $p(1 - \gamma q) : 2\gamma pq : q(1 - \gamma p)$ , где  $\gamma$  – параметр неравновесности популяции, означающий превалирование одних аллелей над другими, например, вследствие инбридинга. При  $\gamma = 1$  (скрещивание происходит абсолютно случайно) получаем классическое соотношение генотипов  $1AA : 2Aa : 1aa$ .

Эволюционные модели часто являются единственным средством описания и изучения процессов, которые слишком медленны, глобаль-

ны, абстрактны и которые происходили в очень далеком прошлом. Отметим, что это касается не только эволюционной биологии, но и таких наук, как палеонтология и астрономия. Даже простая модель Харди–Вайнберга позволяет нам увидеть суть вещей, которая оставалась бы скрытой для исследователя при других обстоятельствах. Например, при допущении, что аллель *A* является доминантным, эта модель позволяет нам оценить число гомозиготных и гетерозиготных особей в популяции, даже если *AA*- и *Aa*-индивиды фенотипически неразличимы.

### Модели изменчивости признаков

Э. Майер утверждал, что дарвинизм сделал «популяционное мышление» центральным не только для биологии, но и для всего современного естествознания (Мауг, 1982). Если вид состоит из особей с индивидуальной изменчивостью и является некоторой изменчивой сущностью, значит, описание некоего вида – это описание его изменчивости. Каким образом это можно сделать? Детальное перечисление всех индивидов и их характеристик будет обладать исчерпывающей полнотой, но не будет удобным для исследования. Ф. Гальтон и другие ранние исследователи в области биометрии понимали, что распределение многих количественных признаков приблизительно отвечает нормальному распределению, и во многих случаях они могут быть охарактеризованы двумя статистическими показателями: средней арифметической, которая описывает центр распределения, и дисперсией, которая показывает характер распределения индивидуальных значений признака вокруг центра. С помощью этих двух статистических характеристик изменчивости мы можем решить ряд задач, разрешение которых было бы затруднено или вообще невозможно при других обстоятельствах. Например, мы можем легко сравнить две популяции по среднему значению и дисперсии какого-либо численного признака (Bulmer, 2003).

Подобным образом можно анализировать связи между различными признаками. В своих трудах Ч. Дарвин часто использовал слово «корреляция» при рассмотрении ситуаций, когда экстремальные отклонения от нормы одного признака имели тенденцию совпадать с экстре-

мальными отклонениями других признаков. В биометрии были разработаны различные количественные методы, которые использовались не только для оценки корреляций между признаками в популяции в целом, но и также для оценки сходства индивидов в подгруппах, например между родственниками. «Регрессионный» коэффициент статистики получил свое название благодаря тому, что он первоначально применялся в ситуациях, при которых наблюдалась тенденция у родителей с крайними значениями некоторых признаков производить потомство с менее отклоняющимися значениями этих признаков; коэффициент регрессии определял степень, с которой фенотипы потомков «регрессировали (или возвращались)» к популяционной средней (Falconer, Mackay, 1996).

Сходство родителей и потомков является центральной идеей дарвиновского естественного отбора. Но возникает вопрос, почему «смешивание наследственности» – представление Ч. Дарвина и большинства его современников в «доменделевскую эпоху» – не приводит к быстрой элиминации изменчивости при все время «усредняющихся» скрещиваниях. Такая потеря изменчивости могла бы положить конец адаптивной эволюции, так как не может осуществляться процесс эффективной селекции среди идентичных индивидов. Р. Фишер, Дж. Холдейн и С. Райт показали, что менделевские законы наследственности прогнозируют наблюдаемое сходство родитель–потомок (включая регрессию к средней), но без прогрессирующей утраты изменчивости в череде поколений. Их модели менделевского наследования, усложненные увеличением числа рассматриваемых генов и учетом влияния окружающей среды, дали начало количественной генетике. Так, например, концепция наследуемости непосредственно связывает генетические и средовые воздействия со степенью сходства между родителями и потомками по некоторому признаку. Этот подход к моделированию изменчивости оказался весьма продуктивным. Он положен в основу всех сельскохозяйственных программ скрещиваний, применяется к теоретическим проблемам развития, половой дифференциации, в экологической оценке емкости ниш, к видообразованию и многим другим биологическим проблемам.

Модели количественной генетики работают с огромным массивом данных по отдельным признакам и компонентам изменчивости внешней среды путем слияния еще большего числа индивидуальных малых эффектов в некоторые интегральные показатели, характерные для биологии. Такие классические модели предполагают, что фенотипическая изменчивость в популяции должна отражать баланс между силами, которые, с одной стороны, повышают и, с другой, снижают ее уровень. С. Райт был одним из первых, кто продемонстрировал, что случайная изменчивость частот аллелей в ограниченных популяциях может приводить к снижению генетического разнообразия. Модели такого процесса, получившего название «генетический дрейф», показали, что скорость потерь обратно пропорциональна размеру популяции ( $N$ ), а модели, включающие мутирование ( $\mu$ ), выявили, что уровень стандартной изменчивости пропорционален произведению  $N\mu$ . В случае, когда это произведение намного больше 1, будет наблюдаться высокий уровень изменчивости, а когда меньше 1 – то уровень изменчивости будет незначительный. Другими словами, если будет возникать более чем одна новая мутация на поколение, то генетический locus будет вариабельным, несмотря на ограниченный размер популяции.

Позже генетический дрейф был хорошо изучен популяционными генетиками, когда Дж. Кингман и другие осознали, что современный генетический материал располагается на терминальных точках ветвей генеалогического дерева, которое, в свою очередь, берет начало от единственной предковой копии (Kingman, 2000). Существуют простые связи между размером популяции и предполагаемой глубиной точек ветвления в «генном дереве», в котором мутации являются случайными маркерами его ветвей. Классические представления о мутационном процессе и дрейфе легко получают новое звучание в этой генеалогической модели, которая была использована для разрешения таких вопросов, как рекомбинация, селекция в различных видах субструктурированных популяций (Hudson, 1990).

Эти простые взаимоотношения, полученные при моделировании процессов в популяциях, могут показаться интуитивно понятными, од-

нако, априорно догадаться о них было сложно. Финальная интегральная теория революционизировала анализ генетической изменчивости внутри популяций, трансформируя и проясняя предмет, который до этого представлялся сверхсложным для понимания. В последнее время данные теории приобретают большое значение как ключевые понятия для анализа межпопуляционной и межвидовой изменчивости последовательностей ДНК.

### Модели эволюции

Простейшие модели естественного отбора непосредственно следуют из расширений закона Харди–Вайнберга. Вместо предположения о равной вероятности выживания и репродукции индивидов всех трех генотипов ( $AA$ ,  $Aa$  и  $aa$ ) мы приписываем им различную жизнеспособность, выражаемую числами, пропорциональными их среднему вкладу в следующее поколение. Используя эти допущения, легко получить формулу, которая дает частоту гена в следующем поколении как функцию частоты гена в предыдущем поколении. Эта связь может быть представлена многими различными способами в зависимости от их трактовки процесса селекции. Например, С. Райтом было показано, что изменение частоты аллеля ( $\Delta p$ ) подчиняется уравнению:

$$\Delta p = \frac{pq}{\bar{W}} \frac{\partial \bar{W}}{\partial p}.$$

Здесь  $\bar{W}$  – средняя жизнеспособность всех индивидов в популяции и  $\partial \bar{W} / \partial p$  – производная средней жизнеспособности по  $p$ . Частота аллеля возрастает ( $\Delta p > 0$ ), если он приводит к увеличению средней жизнеспособности. Это можно выразить метафорой об эволюции как процессе, максимизирующем жизнеспособность, как некоем «восхождении на гору в адаптивном ландшафте».

Даже в своих простейших формах эти базовые селекционные модели дают неожиданные и важные результаты. Например, отбор против вредных рецессивных аллелей оказывается безрезультатным, потому как большинство из них, становясь редкими, встречаются исключительно в гетерозиготах, в которых они имеют жизнеспособность, сравнимую с доминантным аллелем. Таким образом, рецессивные «вредные

аллели» могут сохраняться с поразительно высокими частотами в аутбредных популяциях. Подобные модели могут быть использованы при анализе почти любой биологической ситуации, которая может быть представлена в виде количественных отношений между генотипами, фенотипами и жизнеспособностью.

Влияние отбора также может быть легко добавлено к моделям непрерывной количественной изменчивости при определении изменения среднего значения фенотипа как функции силы отбора и наследуемости признака. Эти количественные генетические модели также могут хорошо предсказывать скорость эволюции, что, в частности, используется в экспериментах с сельскохозяйственно значимыми культурами.

Р. Фишер предположил, что жизнеспособность сама по себе является количественным признаком, находящимся под действием отбора, и скорость ее эволюции пропорциональна имеющемуся разнообразию жизнеспособности (Fisher, 1930). Так как отбор стремится увеличить частоты благоприятных по жизнеспособности аллелей, то вариабельность по жизнеспособности должна снижаться по мере того как частота благоприятных аллелей возрастает, тем самым вытесняя менее благоприятные аллели, и как результат происходит замедление эволюционного процесса. Признаки, тесно сцепленные с жизнеспособностью, также будут иметь сравнительно низкие уровни наследственной изменчивости. Однако почему же тогда сохраняются такие высокие различия в жизнеспособности? Являются ли мутации главным их источником? Возможно, что окружающая среда постоянно изменяется и тем самым изменяется сравнительная жизнеспособность общих и редких аллелей? Эти вопросы нашли отражение в развитии многих современных направлений эволюционных исследований.

Селективно нейтральные мутации под действием генетического дрейфа могут иметь достаточно высокие частоты, замещая все другие аллели в ограниченных популяциях. Интересно, что скорость их фиксации может быть даже равна скорости, с которой они возникают. М. Кимура предложил, что основная эволюция на молекулярном уровне может быть именно такого типа. «Почти нейтральные» мутации (т. е. те, которые понижают жизнеспособность не более чем на

обратную величину к размеру популяции –  $1/N$ ) могут также фиксироваться, хотя и с меньшей скоростью. Таким образом, будет ли мутация эффективно нейтральной, почти нейтральной или подвержена отбору – будет зависеть от размера популяции (Kimura, 1969; Crow, Kimura, 1970; Колчанов, Шиндялов, 1991; Crow, 2008). Как предполагается, малые популяции фиксируют многие мутации с незначительно вредным эффектом, при этом потенциально деградируя. Вредные мутации могут также создавать заметный «генетический груз» в больших популяциях, хотя каждая в отдельности мутация не достигает высокой концентрации.

Модели, описанные выше, рассматривают виды как популяции индивидов с общим «генетическим пулом». Разделение тесно связанного генетического пула, например, при образовании барьеров, затрудняющих миграцию между частями видового ареала, может приводить к обособлению популяций и даже образованию новых видов. Анализ миграции между частично изолированными популяциями показал, что даже единственный мигрант на поколение достаточен для поддержания частот аллелей в субпопуляциях одинаковыми и, таким образом, препятствовать фенотипической и экологической дивергенции, даже при наличии жесткого отбора. Поэтому теоретические исследования процессов видообразования подразумевают почти полную репродуктивную изоляцию в начале процесса видообразования.

Эволюционные связи видов и высших таксонов реализуются через историю видообразования. Мы же видим эту историю ретроспективно, рассматривая ее сквозь анализ последовательностей ДНК современных организмов. Мощные компьютерные программные пакеты, основанные на сложных моделях эволюции последовательностей ДНК, используются для воссоздания всевозможных сценариев истории видообразования и поиска данных, наиболее полно объясняющих современную картину. Реалистичность и сложность сегодняшних эволюционных моделей намного превосходят первые модели времен Дарвина, однако объем накопленных данных и имеющиеся вычислительные ресурсы пока не позволяют говорить о возможности полного описания всего существующего биоразнообразия.



### Модели адаптации

Ч. Дарвин в своих работах уделял много времени размышлениям об эволюции адаптаций. Главная цель его труда «Происхождение видов ...» состояла в том, чтобы показать, что естественный отбор сравнительно «благоприятных» вариантов может привести ко всем наблюдаемым в природе адаптациям. Это утверждение было воспринято лишь небольшим числом современников Ч. Дарвина и подвергалось серьезной критике, отчасти от того, что существенный элемент его теории – наследственная изменчивость – оставался так слабо разработан, что не мог быть смоделирован математически. Это, по-видимому, служит объяснением, почему две трети XX в. были посвящены изучению общих вопросов о ходе эволюции, а вопросам, которые были подняты Ч. Дарвином, уделялось гораздо меньшее внимание.

Только в последней трети прошедшего века, когда на основные вопросы были даны ясные ответы, начался бум эволюционно-экологических исследований, которые имплицитно допускали (как это делал Ч. Дарвин), что если признак отбирается в определенном направлении, то он и эволюционирует в этом направлении (Smith, 1974; Vulmer, 1994). Признаки, изучаемые таким образом, являются типичными «компромиссными признаками», т. е. «хорошие» признаки в конечном итоге могут заместить «лучшие». Например, птицы могут нести больше яиц в стремлении вывести больше птенцов, но такая стратегия может истощить их и тем самым отрицательно сказаться на собственной жизнеспособности. Поэтому максимизация только количества яиц в кладке не приведет в эволюционном плане к лучшему результату. С целью найти «оптимальный» размер кладки (которая бы максимизировала жизнеспособность родителей и потомства) при данном наборе состояний были разработаны модели, описывающие выживание взрослых особей в зависимости от числа яиц. Генетические знания придают сущность этим моделям, однако практический анализ строится в основном на изучении фенотипов.

Гены фигурируют более явно в моделях эволюции полового поведения, половых различий, взаимодействий между родственниками,

совокупная жизнеспособность которых может зависеть от взаимного репродуктивного успеха (Charnov, 1982; Hamilton, 1996). В этих моделях ход эволюции не может быть предсказан путем максимизации средней популяционной жизнеспособности, так как игроки «предпочитают» различные цели, и непримиримые «генетические конфликты интересов» возникают между индивидуумами и даже между генами внутри геномов (Frank, 1998).

Прямой конфликт за ресурсы может привести к частотно-зависимым результатам. Например, индивидуумы, которые обостряют конфликты до высшей степени опасности и агрессии (условно назовем их «ястребами»), могут иметь успех, когда большинство их противников – миролюбивые «голуби»; однако же они будут гибнуть, если большинство их противников будут также «ястребами». Для изучения таких ситуаций была разработана эволюционная теория игр (Smith, 1982). Она показывает, что зачастую не существует оптимальной стратегии, которой индивиды должны придерживаться безотносительно к поведению других членов группы; существует «типичная эволюционно стабильная стратегия», которая является лучшей в том смысле, что она не может быть вытеснена другими стратегиями, если большинство будет придерживаться ее.

77 лет назад Дж. Холдейн в своей книге «The Causes of Evolution» (Haldane, 1932) писал, что проникновение математики в биологию только начинается, и если история науки не есть адекватный путеводитель, то оно продолжится, и эти исследования в совокупности будут представлять начало новой ветви прикладной (био-) математики (биоинформатики). Эта мысль является примером прекрасного предвидения как для его, так и для нашего времени.

Математическая модель может быть стройной и безупречной, однако наложение модели на реальный физический мир с неизбежностью корректирует, если не рушит ее. Достаточно вспомнить ньютоновские законы физики, не работающие на уровнях микро- и макромира. Но, тем не менее, на основе ньютоновских законов человечество получило возможность запускать космические аппараты на околоземную орбиту и к далеким планетам. Теория электромагнитного поля Максвелла позволила

нам пользоваться радио, телевидением и мобильными телефонами, а модели квантовой физики – создать атомную и водородную бомбы, построить атомные электростанции. Возможно, что и математические модели эволюционных процессов позволят человечеству направить их для своего блага и процветания.

Работа поддержана программой фундаментальных исследований Президиума РАН «Биологическое разнообразие» № 23.30 и грантом РФФИ № 09-04-00872-а.

### Литература

- Алтухов Ю.П. Генетические процессы в популяциях. М.: Изд-во ИКЦ «Академкнига», 2003. 431 с.
- Базыкин А.Д., Березовская Ф.С., Буриев Т.И. Динамика системы хищник–жертва с учетом насыщения и конкуренции. Факторы разнообразия в математической экологии и популяционной генетике. Пушкино, 1980. С. 6–33.
- Вольтерра В. Математическая теория борьбы за существование. Пер. с фр. / Под ред. Ю.М. Свиричева. М.: Наука, 1976.
- Дарвин Ч. О происхождении видов путем естественного отбора или сохранении благоприятствуемых пород в борьбе за жизнь. М.; Л.: АН СССР, 1939. 350 с.
- Дейнеко И.В., Кель А.Э., Кель-Маргулис О.В. и др. Моделирование динамики генных сетей, регулирующих клеточный цикл в клетках млекопитающих // Генетика. 2003. Т. 39. № 9. С. 1285–1292.
- Ефимов В.М., Катохин А.В. Применение неметрического многомерного шкалирования для мультиплатформенной обработки микрочиповых экспрессионных данных // Информ. вестник ВОГиС. 2009. Т. 13. № 1. С. 102–108.
- Животовский Л.А. Интеграция полигенных систем в популяциях. Проблемы анализа комплекса признаков. М.: Наука, 1984. 183 с.
- Кайданов Л.З. Генетика популяций. М.: Высш. шк., 1996. 320 с.
- Кель А.Э., Колчанов Н.А., Соловьев В.В. Математическое моделирование в молекулярной биологии и генетике. Теория мутационного процесса: делеции и дубликации. Новосибирск: НГУ, 1989. 86 с.
- Клайн М. Математика. Поиск истины. М.: Мир, 1988. 295 с.
- Клини С. Математическая логика. М.: Мир, 1973. 480 с.
- Колчанов Н.А., Шиндялов И.Н. Теоретическое исследование эволюции регуляторных контуров при различных типах отбора // Проблемы генетики и теории эволюции. Новосибирск: Наука, 1991. С. 268–280.
- Лихошвай В.А., Матушкин Ю.Г., Фадеев С.И. Задачи теории функционирования генных сетей // Сиб. журн. индустр. математики. 2003. Т. 6. № 2. С. 64–80.
- Математические модели эволюционной генетики. Новосибирск: ИЦиГ СО АН СССР, 1980. 160 с.
- Никоро З.С., Журкевич Н.М., Фомичева И.И. Применение формулы Гарди–Вейнберга при анализе генетического полиморфизма // Математические модели эволюционной генетики. Новосибирск: ИЦиГ СО АН СССР, 1980. С. 127–141.
- Ортега-и-Гассет Х. Что такое философия? М.; Новосибирск: Наука, 1991. 404 с.
- Ратнер В.А. Математическая популяционная генетика. Новосибирск: Наука, 1977. 126 с.
- Ратнер В.А., Кананян Г.Х. Построение филогенетических древ для функциональных центров глобинов // Исследования по математической генетике / Ред. В.А. Ратнер. Новосибирск: ИЦиГ СО АН СССР, 1975. С. 125–168.
- Ратнер В.А., Шамин В.В. Сайзеры: моделирование фундаментальных особенностей молекулярно-биологической организации // Математические модели эволюционной генетики. Новосибирск: ИЦиГ СО АН СССР, 1980. С. 66–91.
- Рогозин И.Б., Вульф Ю.И., Бабенко В.Н., Кунин Е.В. Эволюция геномов эукариот и принцип максимальной парсимонии // Информ. вестник ВОГиС. 2005. Т. 9. № 2. С. 141–152.
- Родин С.Н. Идея коэволюции. Новосибирск: Наука, 1991. 268 с.
- Свиричев Ю.М., Пасеков В.П. Основы математической генетики. М.: Наука, 1982. 512 с.
- Фишер Р.А. Статистические методы для исследователей. М.: Госстатиздат, 1958. 268 с.
- Шедько С.В., Гинатулина Л.К., Мирошниченко И.Л., Немкова Г.А. Филогеография митохондриальной ДНК южной азиатской мальмы *Salvelinus cbrilus* (Pallas, 1814) (Salmoniformes, Salmonidae): опосредованная интрогрессия генов? // Генетика. 2007. Т. 43. № 2. С. 227–239.
- Atlas of Protein Sequence and Structure // National Biomedical Research Foundation / Ed. M.O. Dayhoff. 1972. V. 5 (Supplement 3).
- Bulmer M. Theoretical Evolutionary Ecology. Sunderland, Mass., 1994.
- Bulmer M. Francis Galton: Pioneer of Heredity and Biometry. Johns Hopkins Univ. Press, 2003.
- Cavalli-Sforza L.L., Edwards A.W.F. Analysis of human evolution under random genetic drift. Cold Spring Harbor Symposia on Quantitative Biology, 1964. V. 29. P. 9–20.
- Cavalli-Sforza L.L., Edwards A.W.F. Phylogenetic

- analysis: Models and estimation procedures // *Evolution*. 1967. V. 32. P. 550–570.
- Charnov E.L. *The Theory of Sex Allocation*. Princeton, 1982.
- Crow J.F. Mid-Century controversies in population genetics // *Ann. Rev. of Genetics*. 2008. V. 42. P. 1–18.
- Crow J.F., Kimura M. *An Introduction to Population Genetics Theory*. N.Y., 1970.
- Falconer D.S., Mackay T.F.C. *Introduction to Quantitative Genetics*. 4th ed. Essex, England, 1996.
- Felsenstein J. Statistical inference of phylogenies (with discussion) // *J. Roy. Stat. Soc. A*. 1983. V. 146. P. 246–272.
- Fisher R.A. *The Genetical Theory of Natural Selection*. Oxford, 1930.
- Fitch W.M., Farris J.S. Evolutionary trees with minimum nucleotide replacement from amino acid sequence // *J. Mol. Evol.* 1974. V. 3. P. 263–278.
- Fitch W.M., Margoliash E. Construction of phylogenetic trees // *Science*. 1967. V. 155. P. 279–284.
- Frank S.A. *Foundations of Social Evolution*. Princeton, 1998.
- Gillespie J.H. *A Concise Guide to Population Genetics*. Baltimore, 1998.
- Goodman M., Barnabas J., Matsuda G., Moore G.W. Molecular evolution in the descent of man // *Nature*. 1971. V. 233. P. 604–613.
- Green R.E., Krause J., Ptak S.E. *et al.* Analysis of one million base pairs of Neanderthal DNA // *Nature*. 2006. V. 444. P. 330–336.
- Haldane J.B.S. *The Causes of Evolution*. N.Y., 1932.
- Hamilton W.D. *Narrow Roads of Gene Land*. Oxford, 1996.
- Harvey P.H., Pagel M. *The Comparative Method in Evolutionary Biology*. Oxford, 1991.
- Hinnig W. *Phylogenetic Systematics* / Translated by Davis D.D., Zangerl. Urbana, 1966.
- Hudson R.R. Gene Genealogies and the Coalescent Process // *Oxford Surveys in Evolutionary Biology*. 1990. V. 7. P. 1–44.
- Hull D. *Science as a Process*. Chicago, 1988.
- Kimura M. The rate of molecular evolution considered from the standpoint of population genetics // *Proc. Natl Acad. Sci. USA*. 1969. V. 63. P. 1182–1188.
- Kimura M., Ohta T. On some principles governing molecular evolution // *Proc. Natl Acad. Sci. USA*. 1974. V. 71. № 7. P. 2848–2852.
- Kingman J.F.C. Origins of the coalescent: 1974–1982 // *Genetics*. 2000. V. 156. P. 1461–1463.
- Kitching I.J., Forey P.L., Humphries J.C., Williams D.M. *Cladistics: The Theory and Practice of Parsimony Analysis*. Oxford, Oxford Univ. Press, 1998.
- Larget B., Simon D.L. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees // *Mol. Biol. Evol.* 1999. V. 16. P. 750–759.
- Lewis P.O. Phylogenetic systematics turns over a new leaf // *Trends in Ecol. and Evol.* 2001. V. 16. P. 30–37.
- Lotka A.J. *Elements of Physical Biology*. Baltimore: Williams and Wilkins, 1925.
- Lutzoni F., Pagel M., Reeb V. Major fungal lineages derived from Lichen-symbiotic ancestors // *Nature*. 2001. V. 411. P. 937–940.
- Mayr E. *The Growth of Biological Thought*. Cambridge, Mass., 1982.
- Omland K.E., Cook L.G., Crisp M.D. Tree thinking for all biology: the problem with reading phylogenies as ladders of progress // *BioEssays*. 2008. V. 30. P. 854–867.
- Pagel M. Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters // *Proc. Roy. Soc. (B)*. 1994. V. 255. P. 37–45.
- Provine W.B. *The Origins of Theoretical Population Genetics*. Chicago, 1971.
- Ridley M. *Evolution and Classification: The Reformation of Cladism*. Harlow, England, 1986.
- Seeger J., Adler F.R. *Mathematical models* // *Encyclopedia of Evolution* / Ed. M. Pagel. Oxford Univ. Press, 2002. V. 2. P. 684–688.
- Smith M.J. *Mathematical Ideas in Biology*. Cambridge, 1968.
- Smith M.J. *Models in Ecology*. Cambridge, 1974.
- Smith M.J. *Evolution and the Theory of Games*. Cambridge, 1982.
- Sokal R.R., Sneath P.H.A. *Principles of Numerical Taxonomy*. San Francisco, 1963.
- Swofford D.L., Olsen G.J., Waddell P.J., Hillis D.M. *Phylogenetic Inference* // *Molecular Systematics*. 2d ed. / Ed. D.M. Hillis, C. Moritz, B.K. Mable. Sunderland, Mass., 1996.
- Takahata N., Satta Y., Klein J. Divergence time and population size in the lineage leading to modern humans // *Theor. Popul. Biol.* 1995. V. 48. P. 198–221.
- Waterman M.S., Smith T.F., Singh M., Beyer W.A. Additive evolutionary trees // *J. Mol. Biol.* 1977. V. 64. P. 199–213.

## **MODELS IN EVOLUTIONARY BIOLOGY**

**N.N. Yurchenko, I.V. Deineko, I.K. Zakharov**

Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia, e-mail: yurchen@bionet.nsc.ru

### **Summary**

The history of mathematical approaches to evolutionary biology is introduced. General concepts, goals and strategy of mathematical modeling in biology are discussed on the material ranging from original studies of the first biometricians, the contemporaries of Ch. Darwin, through mathematical formalization of microevolution (mutation, migration, isolation, genetic drift) to modern phylogeny reconstructions based on nucleotide sequencing data.